Peer-to-Peer Filesharing and the Market for Digital Information Goods*

Ramon Casadesus-Masanell[†] Andres Hervas-Drane[‡]

May 8, 2006

Abstract

Existing models of peer-to-peer (p2p) filesharing assume that individuals are concerned with each others' wellbeing. Without social preferences (*i.e.*, altruism or reciprocity), peers are better off freeriding whenever the cost of sharing content is larger than that of not sharing. In the absence of social preferences this public-goods problem results in the collapse of the p2p network. Because p2p networks are composed of millions of individuals who interact anonymously, we find inadequate the assumption that peers care about each others' utility. We present microfoundations for a stylized model of a p2p network where all peers are endowed with standard preferences and show that the resulting endogenous structure of the p2p network is conducive to sharing content by a significant number of peers, even if sharing is costlier than freeriding. Selfish utility-maximizing peers are better off sharing because by doing so they face less congestion. We characterize the endogenous level of sharing and present comparative statics results. We build on this framework to analyze the optimal strategy of a profit-maximizing firm, such as Apple's i-Tunes, that offers the same content available on the network. Contrary to the p2p network, the firm offers downloads on a traditional client-server architecture and sells content at positive prices. We show that the firm may be better off setting high prices, allowing the network to survive, and that the p2p network may work more efficiently in the presence of the firm than in its absence.

1 Introduction

Peers in peer-to-peer (p2p) networks face a fundamental choice between sharing content or freeriding. When a peer decides to share content –a costly activity– she is effectively supplying two different goods. On the one hand, she provides *content*. Obviously, the peer who shares does not benefit from the content that she is sharing as she already owns it. On the other hand she also supplies *upload bandwidth* and this may result in lower network congestion. Sharing results in lower congestion if upload bandwidth is a scarce resource. Based on the available empirical evidence, in this paper we assume this to be the case. The nature of peer-to-peer networks warrants that the provision of bandwidth benefits all peers equally in expected terms. In sum, peers face a trade off: by sharing they bear costs that could be avoided by freeriding, but sharing also reduces average network congestion and this benefits every peer, including the peer who shares.

Building on this insight, we construct a model where peers provide bandwidth in addition to content when they decide to share. Specifically, we consider a finite population of agents that derive positive and homogenous utility from digital content. Peers suffer disutility from the costs associated with downloading content. These costs are proportional to the time required to complete downloads, the level of *congestion*, which in turn depends on the bandwidth provision available in the network. Peers may reduce their expected congestion by providing upload bandwidth to other peers. We model this decision as a binary choice: share content or freeride. By having agents differ in their disutility of congestion (impatience or opportunity cost of time) we show that an endogenous level of sharing emerges in the network. Selfish utility-maximizing peers are better off sharing because by doing so they face less congestion. To the best of our knowledge, there is no earlier model of p2p filesharing with endogenous congestion where peers concerned solely about the impact of their actions on their own utility decide to share content.

We build on this framework to analyze the optimal strategy of a profit-maximizing firm that offers the same content available on the network at positive prices employing a traditional client-server architecture. In the absence of altruism towards artists, it is an interesting question why consumers pay to purchase licensed content online. Towards answering this question, we derive the shape of the demand function the firm faces and characterize its optimal

*This abridged draft was prepared for the 2006 Workshop on the Economics of Networked Systems (NetEcon06). An expanded version is available from the authors.
We thank seminar participants at the 2005 Bosphorus Workshop on Economics Design, the 2005 CoCombine conference on the Economics of the Internet, London Business School, Harvard Business School and MIT. We are grateful for financial support from the HBS Division of Research, the Fundación BBVA and the FPU program of the Spanish Ministry of Education and Science.

[†]Harvard Business School. (casadesus@gmail.com)
[‡]Universitat Autònoma de Barcelona. (andresonline@gmail.com)

pricing strategy. In essence, our framework points to the central role of 'convenience' when accessing and consuming digital content through the Internet.

The model captures important stylized facts identified by the literature. First, Asvanund et al. [2] show that congestion worsens with size as peer-to-peer networks grow. Our model endogenously generates this result. In fact, the effect of network size on congestion helps explain the coexistence of multiple different p2p networks. Second, many studies have shown that heavy users of p2p filesharing networks are more prone to purchase content online. Our framework not only suggests that there is no contradiction in this observed behavior, but also sheds light on the factors that explain the demand for online content in the presence of a p2p network. Third, we provide insights on content pricing and the effectiveness of industry initiatives such as suing heavy sharers. Finally, our model shows that filesharing networks strictly benefit from bandwidth infrastructure improvements. This suggests that filesharing is indeed a driver for broadband demand and helps explain why Internet service providers have not taken action to limit the spread of p2p applications and filesharing traffic load. We believe that our results should be of interest to all participants in markets for digital information goods.

The paper is organized as follows. Section 2 introduces the building blocks of our model of peer-to-peer filesharing and describes the game (in the absence of a profit maximizing firm). In Section 3 we present a simple approximation to the average congestion in an arbitrary peer-to-peer network. Section 4 derives the equilibrium network configuration and studies its properties. In Section 5 we further characterize the equilibrium under the assumption that peers' time preferences are independently drawn from a uniform probability distribution. Finally, in Section 6 we introduce a profit maximizing firm that competes against the p2p network and analyze the interdependencies that arise in the competition between both business models.

2 The model

We consider a population of M agents that derive utility from the consumption of digital information goods. They all value content equally and differ only in their disutility of congestion. We model the formation of a peer-to-peer network in two stages. In the first stage, agents choose (simultaneously) whether or not to join the network. Agents who choose to belong to the network can either share their content or freeride. Sharers offer their content on the network for download by other peers while freeriders do not. While sharing content is costly, some sharing is required for the network not to collapse as downloads can only be realized from other sharers. We will refer to agents in the network as *peers* and those outside as *outsiders*. We let $N \leq M$ denote the number of peers. $M - N$ is the number of outsiders.

In the second stage peers interconnect and downloads are realized. The utility of a peer that freerides is given by

$$u_i^f = u_d - (c_n + \rho_i)t_d, \tag{1}$$

and that of a peer who shares his content is

$$u_i^s = u_d - (c_n + c_s + \rho_i)t_d, \tag{2}$$

where $i \in \mathbf{N} = \{1, 2, ..., N\}$. Outside utility is normalized to zero.

The utility derived from content once a download has been completed is u_d and it is common across all agents. The time required to complete a download, t_d, is endogenous and depends on the level of congestion. A lower bandwidth transmission speed implies higher level of congestion resulting in higher download time. Every peer suffers a positive cost c_n of pertaining to the network. This captures the costs of the computing resources and the bandwidth for signalling traffic required to remain connected to the network until a download completes. Sharers additionally bear cost c_s. This is the cost originating from offering content for download on a public p2p network (including expected costs of legal action against the peer) as additional computing resources (storage space) and upload bandwidth is required.

Parameter $\rho_i \geq 0$ reflects the disutility of congestion experienced by peer i. The larger ρ_i is, the higher the disutility the peer obtains from an increase in the time required to complete a download. Hence ρ_i can alternatively be interpreted as impatience or opportunity cost of time: how much peer i values quick accessibility to content. Without loss of generality we choose indexes i so that $\rho_i \leq \rho_{i+1}$ for all i. All other costs being equal, peers would prefer to obtain the downloadable content immediately avoiding congestion delays. An increase in the time required to complete a download reduces the utility obtained from the network by increasing both the network costs and the disutility of congestion experienced by all peers.

To solve the second stage we let $\mathbf{S} \subset \mathbf{N}$ be the set of sharers in the network (given the agents' first-stage strategies) and denote by S the number of sharers (the cardinality of $\mathbf{S}$). A downloader exclusively served by a sharer will download a unit of content in time $\theta > 0$; that is, $t_d = \theta$. This can be interpreted as θ capturing the relation between the filesize of content and the bandwidth capacity available to peers. Thus an improvement in either encoding efficiency reducing filesizes or broadband infrastructure increasing bandwidth amounts to a reduction in θ. Download bandwidth is assumed not to be a limiting factor. If more than one downloader is connected to a given sharer, bandwidth is shared evenly amongst them. This can be interpreted as downloading taking place simultaneously or, alternatively, the sharer serving download queues for fractions of content by turns.

A set of links connecting peers to sharers where every peer connects to one sharer only and no sharer connects to herself is called a *network allocation*. A *stable* network allocation is one where no peer can be made strictly better off by connecting to a different sharer. We assume that following the first stage (where peers decide whether to share or to freeride) a stable network allocation ensues. Clearly, if the network allocation was not stable, at least one peer would have an incentive to connect to a different sharer.

The following mild assumption is required for the results: $u_d > (c_n + c_s + \rho_i)\theta$ for all i. This ensures that a p2p network with minimum congestion is always preferred to the outside option of not pertaining to the network. With the notation in place, we now proceed to solving the game by backwards induction.

3 Network foundation

Interconnection occurs in the second stage, after each peer has decided whether she will share or freeride. Congestion plays a crucial role in our development as peers choose to share taking into consideration the effect that their sharing has on congestion. Given a network allocation, the bandwidth obtained by peer $i \in \mathbf{N}$ can be computed as follows: if the peer is connected to a sharer to which k other peers are connected to, then peer i obtains effective bandwidth $1/(k+1)$.

Freeriders can connect to every sharer and, thus, have S possible links available to choose from. Sharers, on the other hand, cannot connect to themselves. As a consequence, sharers have $S-1$ possible links available. This implies that, in general, the expected congestion of sharers and freeriders will differ. To compute the *expected bandwidth* for freeriders and sharers in a network with N peers and S sharers, we begin by computing each peer's effective bandwidth in every stable network allocation. We then average these effective bandwidths assuming that every stable network allocation is equally likely. *Expected congestion*, the delay required to complete downloads, t_d, is the inverse of the expected bandwidth.

In another paper [3] we derive an exact expression for the expected effective bandwidth of sharers and freeriders. There, we show that S/N is a good approximation to the expected bandwidth of both sharers and freeriders. The accuracy of this approximation increases with the size of the network. In fact, already in a network of size $N = 10$, the expected effective bandwidth of sharers and freeriders differs from S/N by, at most, 0.0012. Given this result, we conclude that all peers obtain an expected download bandwidth close to S/N. This implies that the expected time to complete a download for all peers can be approximated by $t_d = \theta / \frac{S}{N} = \theta \frac{N}{S}$. It should be noted that although the expected bandwidth depends linearly in the number of sharers, the time required to complete a download does not. This property is crucial to our results. Technically, it ensures that our objective function is concave in S, allowing for interior equilibria in which sharing and freeriding may coexist for certain ranges of N.

4 Equilibrium network configurations

In this section we analyze the first stage of the game. Every peer i chooses whether to freeride or to share content (at additional cost c_s). In making their decision, peers consider the effect of their choice on expected download time $\theta \frac{N}{S}$. Equations (1) and (2) imply that if expected download time was *not* affected by the sharing decision, no peer would ever share and the peer-to-peer network would not be viable.

In this section we take N as given. This amounts to assuming that all N peers in the network obtain positive utility. In general, this will depend on S and the distribution of ρs. In the following section we relax this assumption and let peers decide whether or not to join the network.

Let $P = \{\mathbf{F}, \mathbf{S}\}$ be a partition of $\mathbf{N}$. We refer to P as a *network configuration*.[1] $\mathbf{F}$ is the set of freeriders and $\mathbf{S}$ the set of sharers. Obviously, P constitutes a Nash equilibrium if no $i \in \mathbf{S}$ prefers to (unilaterally) become a freerider and no $j \in \mathbf{F}$ prefers to become a sharer.

Proposition 1 *Every equilibrium network configuration $P = \{\mathbf{F}, \mathbf{S}\}$ has the following form: $\mathbf{F} = \{1, 2, ..., n-1\}$ and $\mathbf{S} = \{n, n+1, ..., N\}$ for some $n \in \mathbf{N}$. The system of equations given by Γ_s identifies the set $\mathbf{S}$ for all equilibrium network configurations,*

$$\Gamma_s = \{ i \in \mathbf{I} \,|\, H(\rho_i) \subset G(\rho_i) \},$$

where

$$\begin{aligned} G(\rho_i) &= \left\{ k \in \mathbf{I} \,\middle|\, \frac{c_f + \rho_{i-1}}{c_s} \le k \le \frac{c_f + c_s + \rho_i}{c_s} \right\} \\ H(\rho_i) &= N + 1 - i. \end{aligned}$$

Proof. All proofs are in the appendix. ■

The proposition says that if peer i is a sharer in equilibrium network configuration P, then peer $i+1$ must also be a sharer. Moreover, if peer j is a freerider, then peer $j-1$ must also be a freerider. Thus, the most impatient peers prefer to share while

[1] Notice that a network configuration can be mapped to many different network allocations.

the more patient peers are better off freeriding. The reason is simple: by sharing content, peers reduce congestion and the (positive) marginal effect on peer utility implied by lower congestion is proportional to the value of ρ_i. Peers for whom the opportunity cost of time is high, are more inclined to share. This is true even though given any *fixed* level of congestion, all peers (regardless of the value of ρ) are better off freeriding than sharing.

The system of equations $\{G(), H()\}$ characterizes the equilibrium network configurations by pinning down to the fullest possible extent the set of sharers $\mathbf{S}$. Note that certain parameter constellations may exhibit multiple equilibria and Γ_s may not be a singleton.

Let $\mathbf{S} = \{n, n+1, ..., N\}$ be the set of sharers in an equilibrium network configuration. We refer to the case $n = 1$ as a *full-sharing network configuration* (or *full-sharing equilibrium*) and to the case $n > 1$ as a *partial-sharing network configuration* (or *partial-sharing equilibrium*). In a full-sharing network configuration all peers are sharers. In this case, congestion is minimized as the expected download time for all peers (t_d) is equal to θ.

Remark 2 *Full-sharing holds in the network if and only if*

$$N < \frac{c_n + c_s + \rho_1}{c_s}.$$

Therefore, if N is sufficiently small, the unique equilibrium network configuration has all peers sharing content. Notice that as the incremental cost of sharing c_s approaches zero, the maximal network size that supports full sharing grows without bound. When N is large, the equilibrium network configurations will typically entail partial sharing. In this case, expected download time will be larger than θ for all peers.

5 Equilibrium with $\rho_i \sim U[0, \bar{\rho}]$

In Section 4 we have characterized all equilibrium network configurations for the general case, without specific assumptions on the distribution of ρ_is or the cardinality of $\mathbf{N}$. In order to ensure tractability when we introduce a profit maximizing firm (Section 6), we make the additional assumption that ρ_is are i.i.d. $U[0, \bar{\rho}]$. This allows us to further characterize the set of equilibrium network configurations.

The first result shows that if the network has many peers, then the set of equilibrium configurations is a singleton.

Remark 3 *For N large enough, there is a unique equilibrium network configuration.*

The next result identifies the most patient sharer as a function of the parameters. Identifying precisely the most patient sharer will allow us to easily analyze how the different parameters affect network congestion. In particular, we are interested on the effect that N has on congestion. If congestion decreases when N grows, then the p2p network becomes gradually more valuable as the number of peers expands. If, in contrast, network congestion grows with N, then the p2p network exhibits negative (network) externalities.

Proposition 4 *Let $\rho_{s(N)}$ be the most patient sharer in equilibrium. Then, for N large,*

$$\rho_{s(N)} \simeq \frac{\bar{\rho}\left((N-1)\, c_s - c_n\right)}{\bar{\rho} + N c_s},$$

and the cardinality of $\mathbf{S}$ *in equilibrium is given by*

$$S = \frac{\bar{\rho} + c_s + c_n}{\frac{1}{N}\bar{\rho} + c_s}.$$

Notice that $\rho_{s(N)}$ is increasing in N. This implies that the larger the cardinality of $\mathbf{N}$, the lower is the proportion of sharers in equilibrium. In other words, in our model the p2p network exhibits negative network externalities (past the threshold network size of full sharing): the larger the number of peers in the network, the lower the average utility that peers obtain. In fact, as $N \to \infty$, $\rho_{s(N)} \to \bar{\rho}$. Therefore, when the network is very large, only the most impatient peer winds up sharing content. Also note that $S < \infty$, even as $N \to \infty$. Therefore, not only the proportion of sharers dwindles as N grows, but the absolute number of sharers has a cap. As a consequence, the expected download time for all peers ($\theta \frac{N}{S}$) grows without bound as N increases. This means that as N grows, the peer-to-peer network becomes less and less attractive. We will now see that this has important implications for the equilibrium pricing strategy of a profit maximizing firm competing for customers against a p2p network.

6 The firm

We next introduce the problem of an online firm selling digital information goods also available on the peer-to-peer network. To the firm, the network is a competitor because peers that choose to download files from the network could otherwise become paying customers. Because content is free on the p2p network, for the firm to persuade users of digital content to purchase it at positive price, it must offer added benefits that the p2p network cannot match. In our view, the most important advantage of the firm is that it can offer lower download time than the network. Specifically, we let the firm offer content streaming based on traditional client-server architecture. That is, consumption of content acquired through the firm can be realized immediately at the moment of purchase. As a consequence, the utility of buyers is:

$$u_i = u_d - p. \tag{3}$$

Notice that (3) is the natural adaptation of (1) and (2) to the case of streaming. With streaming, the expected download time t_d falls down to zero. Thus, the terms $c_n + \rho_i$ and c_s do not appear in (3). On the other hand, the firm charges a positive price for content. We assume that the content offered by the firm is the same as that shared in the p2p network.

Agents may now choose to purchase content instead of downloading it off the network at zero price. We modify the timing of the game accordingly. In the first stage, the firm chooses the price at which content is sold. In the second stage, agents choose to either purchase from the firm, enter the network, or stay outside and not consume content. Agents who enter the network may share or freeride. In the third stage, agents in the network interconnect and downloads take place.

We assume that the firm faces zero marginal costs. All infrastructure and running costs related to the service are fixed and independent of the activity level. The assumption captures the fact that selling additional copies of digital content has negligible incremental costs.

The problem of the firm is to quote the price p that maximizes profits. The following proposition summarizes the firm's optimal strategy as a function of the parameters.

Proposition 5 *Let*

$p_{fc} := \theta(c_n + c_s)$	*(full market coverage)*
$p_{hc} := \frac{1}{2}\theta(\bar{\rho} + c_n + c_s)$	*(high market coverage)*
$p_{lc} := \frac{1}{2}\theta(\bar{\rho} + Mc_s)$	*(low market coverage)*
$p_{oc} := u_d$	*(outsiders only coverage)*

The optimal pricing strategy is given by:

- *If* $\bar{\rho} > c_n + c_s$, *then*

$$\begin{array}{lll} p_{hc} & \textit{if} & M < M_b \\ p_{lc} & \textit{if} & M_b \leq M < M_c \\ p_{oc} & \textit{if} & M \geq M_c \end{array}$$

- *If* $\bar{\rho} \leq c_n + c_s$, *then*
  - *If* $u_d \leq 2\theta(c_n + c_s)$, *then*

$$\begin{array}{lll} p_{fc} & \textit{if} & M < M_d \\ p_{oc} & \textit{if} & M \geq M_d \end{array}$$

  - *If* $u_d > 2\theta(c_n + c_s)$, *then*

$$\begin{array}{lll} p_{hc} & \textit{if} & M < M_b \\ p_{lc} & \textit{if} & M_b \leq M < M_c \\ p_{oc} & \textit{if} & M \geq M_c \end{array}$$

where $M_a = \frac{4(c_n+c_s)-\bar{\rho}}{c_s}$, $M_b = \frac{(c_n+c_s)(2\bar{\rho}+c_n+c_s)}{\bar{\rho}c_s}$, $M_c = \frac{2u_d-\theta\bar{\rho}}{\theta c_s}$ *and* $M_d = \frac{u_d^2-\theta\bar{\rho}(u_d-\theta(c_n+c_s))}{\theta c_s(u_d-\theta(c_n+c_s))}$.

Equilibrium profits are:

$\pi_{fc} = M\theta(c_n + c_s)$	*(full market coverage)*
$\pi_{hc} = \frac{M\theta(\bar{\rho}+c_n+c_s)^2}{4\bar{\rho}}$	*(high market coverage)*
$\pi_{lc} = \frac{1}{4}M\theta(\bar{\rho} + Mc_s)$	*(low market coverage)*
$\pi_{oc} = u_d(1 - \frac{u_d}{\theta(\bar{\rho}+Mc_s)})M$	*(outsiders only coverage)*

The firm's demand curve is downward sloping as expected. Agents with sufficiently high disutility of congestion prefer to purchase instead of belonging to the network. The lower the price the firm quotes, the higher the number of agents who prefer to purchase. The agents who obtain higher surplus from purchasing are those with larger values of ρ, agents who may have potentially remained outsiders in the absence of the firm. The size of the network is affected by the presence of the firm, as only those individuals with lower disutility of congestion remain as peers; agents who would otherwise share content may now leave the p2p network and purchase from the firm. As sharers leave, peers that might have otherwise been freeriders are now better off sharing.

The demand curve exhibits a non-derivability. Two ranges exist over which congestion in the network differs. In the lower price range, full sharing holds and congestion is not affected by peers entering or exiting to purchase. In the higher price range, only partial sharing holds. In this case congestion varies with the size of the network and this effect is taken into account by agents; the smaller the network, the lower the level of congestion. As an example, consider the effect of a reduction in price. In the full sharing range, peers who switch to purchase are not affecting the congestion experienced by those remaining. But in the case of partial sharing, peers leaving are (indirectly) reducing congestion by reducing the size of the network. This effect ensures that less peers will react to a price reduction in the case of partial sharing.

Proposition 5 shows that the optimal pricing strategy depends critically on market size. The firm will only quote a low price and cover the entire market if it is sufficiently small. The bigger the market, the more profitable it is for the firm to target agents with a higher disutility of congestion by quoting higher prices. If the market is sufficiently large, it is optimal for the firm to serve outsiders exclusively. The intuition for this result lies on the mechanisms that drive congestion in a peer-to-peer network. As the size of the network increases, so does the congestion experienced by all peers. As a consequence, the surplus that the firm can extract by targeting agents who most suffer congestion grows more than that obtained by covering the whole market by quoting a low price. The fact that this result arises under a uniform distribution of ρ suggests that it is quite general.

For the case of a large market, we should expect the firm to quote a high price, close to the actual valuation of content. Such a pricing strategy does indeed internalize the presence of the network. As a result, the firm will not directly affect the size of the network, as agents who purchase would otherwise choose to stay outside, although the profits of the firm would strictly increase if the p2p network did not exist. The strategic effect of the presence of a p2p network on the firm can be likened to a low quality firm competing against a vertically differentiated competitor. The firm has strong incentives to offer high quality service by investing to minimize congestion and to quote a high price. The net effect of the network's presence on consumer welfare is unambiguously positive.[2]

It is of interest to look at the effect on firm's profit of changes in the technology parameter θ and the marginal cost of sharing c_s. A technology improvement captured by a decrease in θ strictly decreases the firm's profit. All other factors equal, a technology improvement in either broadband infrastructure or the efficiency of digital encoding reduces effective download delays in the network making it more attractive. Similarly, profits are increasing in the marginal cost of sharing under all market configurations. Because c_s measures the efficiency of the network, or how well the network scales with size, a higher marginal cost of sharing implies higher levels of congestion. This benefits the firm by generating higher surplus to be extracted out of potential purchasers. These effects provide strong incentives for the firm to intervene. Strategies that may serve this purpose include traffic discrimination on broadband networks, prioritizing the firm's data, and randomly suing sharers thereby increasing the expected cost of sharing for all agents.

References

[1] Antoniadis, P. and Courcoubetis, C. and Mason, R. (2004), 'Comparing Economic Incentives in Peer-to-Peer Networks,' *Computer Networks*, Vol.46, pp.133-146.

[2] Asvanund, A. and Clay, K. and Krishnan, R. and Smith, M. D. (2004), 'An Empirical Analysis of Network Effects in Peer-to-Peer Music-Sharing Networks,' *Information Systems Research*, Vol.15, pp.155-174.

[2]It is important to stress that in our framework the production of content is unaffected by the free exchange of content between peers in the p2p network. If lower royalties imply lower incentive by artists to produce quality content, then the net effect of the presence of the p2p network is ambiguous.

[3] Creus Mir, A. and Casadesus-Masanell, R. and Hervas-Drane, A. (2006), 'Bandwidth Allocation in Peer-to-Peer Filesharing Networks,' Mimeo, Harvard Business School.

[4] Cunningham, B. M. and Alexander, P. J. and Adilov, N. (2004), 'Peer-to-Peer File Sharing Communities,' *Information Economics and Policy*, Vol.16, Issue 2, pp.197-221.

[5] Feldman, M. and Lai, K. and Chuang, J. and Stoica, I., 'Quantifying Disincentives in Peer-to-Peer Networks,' 1st Workshop on Economics of Peer-to-Peer Systems (2003).

[6] Feldman, M. and Papadimitriou, C. and Chuang, J. and Stoica, I. (2004), 'Free-riding and Whitewashing in Peer-to-Peer Systems,' PINS '04: Proceedings of the ACM SIGCOMM workshop on Practice and theory of incentives in networked systems, pp.228-236.

[7] Golle, P. and Leyton-Brown, K. and Mironov, I. and Lillibridge, M. (2001), 'Incentives for Sharing in Peer-to-Peer Networks,' *Lecture Notes in Computer Science*, Vol.2232, pp.75+.

[8] Krishnan, R. and Smith, M. D. and Tang, Z. and Telang, R. 'The Virtual Commons: Understanding Content Provision in Peer-to-Peer File Sharing Networks,' November 2004.

A Appendix

Proof of Proposition 1. Sharer $i \in \mathbf{S}$ will not free ride iff:

$$u_d - (c_f + c_s + \rho_i)\,\theta \frac{N}{S} \geq u_d - (c_f + \rho_i)\,\theta \frac{N}{S-1}, \quad (4)$$

or

$$\frac{S}{S-1} \leq \frac{c_f + \rho_i}{c_f + c_s + \rho_i}.$$

Notice that

$$\frac{d\left(\frac{c_f+\rho_i}{c_f+c_s+\rho_i}\right)}{d\rho_i} = \frac{c_s}{(c_f + c_s + \rho_i)^2} > 0. \quad (5)$$

Therefore, if (4) is satisfied for sharer $i \in \mathbf{S}$ it is also satisfied for all sharers i' with $\rho_{i'} \geq \rho_i$. Thus, the more impatient a sharer is, the less the incentive to become a freerider.

A freerider $j \in \mathbf{F}$ will not want to become a sharer iff:

$$u_d - (c_f + \rho_j)\,\theta \frac{N}{S} \geq u_d - (c_s + \rho_j)\,\theta \frac{N}{S+1}, \quad (6)$$

or

$$\frac{S}{S+1} \geq \frac{c_f + \rho_j}{c_f + c_s + \rho_j}.$$

Notice that (5) implies that if (6) is satisfied for peer $j \in \mathbf{F}$ it is also satisfied for all peers $j' \in \mathbf{F}$ with $\rho_{j'} \leq \rho_j$. Thus, the more patient a freeriding peer is, the less the incentive to become a sharer.

We now further characterize the equilibrium network configurations by pinning down to the fullest possible extent the cardinality of S. Let $P = \{\mathbf{F}, \mathbf{S}\}$ be an equilibrium network configuration. Let ρ_i be the most patient sharer in $\mathbf{S}$. Equations (4) and (6) imply that

$$S \leq \frac{c_f + c_s + \rho_i}{c_s} \qquad \text{and} \qquad S \geq \frac{c_f + \rho_{i-1}}{c_s}.$$

Thus,

$$\frac{c_f + \rho_{i-1}}{c_s} \leq S \leq \frac{c_f + c_s + \rho_i}{c_s}. \tag{7}$$

Let $\mathbf{I}$ be the set of integers. The following two objects are useful in what follows:

$$G(\rho_i) = \left\{ k \in \mathbf{I} \left| \frac{c_f + \rho_{i-1}}{c_s} \leq k \leq \frac{c_f + c_s + \rho_i}{c_s} \right. \right\} \tag{8}$$

and

$$H(\rho_i) = N + 1 - i. \tag{9}$$

Correspondence G indicates the cardinality of $\mathbf{S}$ if the sharer with lowest impatience has time preference parameter ρ_i. Function H tells us the number of peers with parameter ρ_j larger than or equal to that of peer i. The solution to the system of equations given by G and H pins down the set of most patient sharers for all equilibrium network configurations:

$$\Gamma_s = \{ i \in \mathbf{I} \, | H(\rho_i) \subset G(\rho_i) \}.$$

Because $G(\rho_i)$ is a correspondence, Γ_s may not be a singleton. Clearly, the cardinality of the set of equilibrium network configurations, coincides with that of Γ_s. ■

Proof of Remark 2. For a full-sharing network configuration to obtain, every peer must realize higher utility sharing than freeriding. In particular, the most patient peer (ρ_1) must be better off sharing than freeriding (given that everybody shares):

$$u_d - (c_n + c_s + \rho_1)\,\theta \geq u_d - (c_n + \rho_1)\,\theta \frac{N}{N-1}.$$

Solving for N we obtain:

$$N \leq \frac{c_n + c_s + \rho_1}{c_s}. \tag{10}$$

■

Sketch of Proof of Remark 3. Recall that the set of equilibrium network configurations is given by

$$\Gamma_s = \{ i \in \mathbf{I} \, | H(\rho_i) \subset G(\rho_i) \}.$$

H is a decreasing function of ρ_i. G is an increasing correspondence. However, when N is large ρ_{i-1} is close to ρ_i. In fact as $N \to \infty$, $|\rho_{i-1} - \rho_i| \to 0$. Thus, when N is large,

$$G(\rho_i) = \left\{ k \in \mathbf{I} \left| \frac{c_f + \rho_i}{c_s} \leq k \leq \frac{c_f + \rho_i}{c_s} + 1 \right. \right\}.$$

Given this, $G(\rho_i)$ is single-valued except at those ρ_i such that $\frac{c_f + \rho_i}{c_s}$ is a natural number. Suppose that at ρ'_i, $\frac{c_f + \rho'_i}{c_s} \in \mathbf{I}$. Then, for all $\varepsilon > 0$, $\frac{c_f + \rho'_i + \varepsilon}{c_s} \notin \mathbf{I}$ and $\frac{c_f + \rho'_i - \varepsilon}{c_s} \notin \mathbf{I}$. Thus, $G(\rho_i)$ is a step function with 'continuous jumps' at $\rho_i \in [0, \bar{\rho}]$ such that $\frac{c_f + \rho_i}{c_s} \in \mathbf{I}$. As a consequence, Γ_s is a singleton. ■

Proof of Proposition 4. We look for $\rho_{s(N)}$ such that the set of peers with $i \geq s(N)$ all want to share. Because $\rho_{s(N)}$ is the most patient sharer, the cardinality of the set of sharers is $S = N - s(N) + 1$.

For $\mathbf{S}$ to be the set of sharers of a stable partition, we need that the most patient sharer does not want to freeride:

$$u_d - \left(c_n + c_s + \rho_{s(N)}\right)\theta \frac{N}{S} \geq u_d - \left(c_n + \rho_{s(N)}\right)\theta \frac{N}{S-1}$$

This expression implies that

$$\rho_{s(N)} \geq (N - s(N))\, c_s - c_n.$$

Therefore, for $S = N - s(N) + 1$ to be stable, $\rho_{s(N)}$ must satisfy $\rho_{s(N)} \geq (N - s(N))\, c_s - c_n$.

We also need that the most impatient freerider does not want to share:

$$u_d - \left(c_n + \rho_{s(N)-1}\right)\theta \frac{N}{S} \geq u_d - \left(c_n + c_s + \rho_{s(N)-1}\right)\theta \frac{N}{S+1}$$

This expression implies that

$$\rho_{s(N)-1} \leq (N - s(N) + 1)\, c_s - c_n.$$

Suppose now that all ρ_is are drawn from a uniform distribution $\rho_i \sim U[0, \bar{\rho}]$. When N is large we have that $s(N) - 1 \simeq \frac{\rho_{s(N)-1}}{\bar{\rho}} N$. Furthermore, large N also implies that $\rho_{s(N)} \simeq \rho_{s(N)-1}$. Therefore $s(N) \simeq \frac{\rho_{s(N)-1}}{\bar{\rho}} N + 1 \simeq \frac{\rho_{s(N)}}{\bar{\rho}} N + 1$. Substituting in the expression above, we obtain

$$\begin{aligned} \left(N - \frac{\rho_{s(N)}}{\bar{\rho}} N - 1\right) c_s - c_n &\leq \rho_{s(N)} \\ \frac{\bar{\rho}\left((N-1)\, c_s - c_n\right)}{\bar{\rho} + N c_s} &\leq \rho_{s(N)}. \end{aligned}$$

When N is large we have that $s(N) - 2 \simeq \frac{\rho_{s(N)-2}}{\bar{\rho}} N$. Furthermore, large N also implies that $\rho_{s(N)-1} \simeq \rho_{s(N)-2}$. Therefore $s(N) - 2 \simeq \frac{\rho_{s(N)-2}}{\bar{\rho}} N \simeq \frac{\rho_{s(N)-1}}{\bar{\rho}} N$ or $-s(N) + 1 \simeq -\frac{\rho_{s(N)-1}}{\bar{\rho}} N - 1$. Now, substituting in the expression above, we obtain

$$\begin{aligned} \rho_{s(N)-1} &\leq (N - s(N) + 1)\, c_s - c_n \\ \rho_{s(N)-1} &\leq \frac{\bar{\rho}\left((N-1)\, c_s - c_n\right)}{\bar{\rho} + N c_s} \end{aligned}$$

So, when N is large we have that

$$\rho_{s(N)-1} \leq \frac{\bar{\rho}\left(\left(N-1\right)c_s - c_n\right)}{\bar{\rho} + Nc_s} \leq \rho_{s(N)}.$$

We conclude that when N is large

$$\rho_{s(N)} \simeq \frac{\bar{\rho}\left(\left(N-1\right)c_s - c_n\right)}{\bar{\rho} + Nc_s}.$$

To identify the cardinality of S, we have that $S = N - s(N) + 1$ and $s(N) - 1 \simeq \frac{\rho_{s(N)-1}}{\bar{\rho}}N$. Therefore,

$$\begin{aligned} S &= N - \frac{\left(N-1\right)c_s - c_n}{\bar{\rho} + Nc_s}N \\ &= N\left(\frac{\bar{\rho} + c_s + c_n}{\bar{\rho} + Nc_s}\right). \end{aligned}$$

■

Proof of Proposition 5. An agent with disutility of congestion ρ_i will only purchase from the firm if:

$$u_d - p \geq u_d - (c_n + c_s + \rho_i)t_d.$$

Because $t_d \geq \theta$ is positive, if the condition is satisfied for peer i it will also be satisfied for peer $i+1$. To solve for demand given a price p we proceed by identifying the indifferent buyer, denoted by ρ_b. If $p = u_d$, only outsiders buy from the firm, as all other agents obtain strictly positive utility in the network. If $p > u_d$ purchasing yields negative utility and the firm faces no demand. To obtain demand when $p \leq u_d$ we must solve for ρ_b, given by:

$$u_d - p = u_d - (c_n + c_s + \rho_b)t_d. \quad (11)$$

Because either full or partial sharing may hold in the network, we consider two separate cases. We begin with the partial sharing case. Substituting $t_d = \theta\frac{N}{S}$ in (11) and taking into account that congestion will depend on ρ_b, as only agents such that $\rho_i \leq \rho_b$ are present in the network:

$$u_d - p \simeq u_d - \left(c_n + c_s + \rho_b^{ps}\right)\theta\frac{N\left(\rho_b^{ps}\right)}{S\left(\rho_b^{ps}\right)},$$

where

$$N\left(\rho_b^{ps}\right) = \frac{\rho_b^{ps}}{\bar{\rho}}M,$$

and

$$S\left(\rho_b^{ps}\right) = N\left(\rho_b^{ps}\right)\left(\frac{\rho_b^{ps} + c_s + c_n}{\rho_b^{ps} + N\left(\rho_b^{ps}\right)c_s}\right).$$

Solving for ρ_b^{ps} yields:

$$\rho_b^{ps} = \frac{p\bar{\rho}}{\theta(\bar{\rho} + Mc_s)}.$$

We next consider the full sharing case and solve for the indifferent buyer by substituting $t_d = \theta$ in (11):

$$u_d - p \simeq u_d - \left(c_n + c_s + \rho_b^{fs}\right)\theta,$$

thus

$$\rho_b^{fs} = \frac{p - \theta(c_n + c_s)}{\theta}.$$

The demand function for the firm is given by:

$$D = (1 - \frac{\rho_b}{\bar{\rho}})M. \quad (12)$$

Substituting ρ_b^{ps} we obtain the expression for demand in the partial sharing range:

$$D^{ps} = (1 - \frac{p}{\theta(\bar{\rho} + Mc_s)})M.$$

And substituting ρ_b^{fs} in (12) we obtain demand in the full sharing range:

$$D^{fs} = (1 - \frac{p - \theta(c_n + c_s)}{\theta\bar{\rho}})M.$$

Full market coverage is obtained when $\rho_b^{fs} = 0$, which implies:

$$p_{fc} = \theta(c_n + c_s).$$

A lower price will also ensure that the market is covered.

We next consider the optimal pricing strategy of the firm. Given that either full or partial sharing may hold in the network, the firm faces two separate cases. Profits in the lower price range, under full sharing, are given by D^{fs}:

$$\pi_{lr} = p(1 - \frac{p - \theta(c_n + c_s)}{\theta\bar{\rho}})M, \quad (13)$$

which has a maximum at

$$p_{hc} = \frac{1}{2}\theta(\bar{\rho} + c_n + c_s).$$

We denote the maximum by p_{hc}, as high coverage of the market is obtained in this price range. In the higher price range, under partial sharing, profits are given by D^{ps}:

$$\pi_{hr} = p(1 - \frac{p}{\theta(\bar{\rho} + Mc_s)})M, \quad (14)$$

which has a maximum at

$$p_{lc} = \frac{1}{2}\theta(\bar{\rho} + Mc_s).$$

As market coverage is lower in this range, we denote the maximum by p_{lc}

To solve the firm's optimal price strategy, profits given by the optimal price in both ranges need to be compared under all feasible parameter configurations. Solving the systems of inequalities implied determines the profit-maximizing price as a function of all the parameters. ■